

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2013

Các Phương Pháp Phân Tích Định Lượng

Bài tập 11

MÔ HÌNH CÓ BIẾN ĐỘC LẬP LÀ BIẾN GIẢ, MÔ HÌNH CÓ BIẾN PHỤ THUỘC LÀ BIẾN NHỊ PHÂN

Ngày Phát: Thứ Năm, 26/12/2013

Ngày Nộp: Thứ Sáu, 03/01/2014

Bản in nộp lúc 8:20 sáng, tại Hộp nộp bài tập trong phòng Lab

Bản điện tử gửi theo hướng dẫn của trợ giảng.

Bài 1

Bạn hãy sử dụng file dữ liệu Employee Data.xls đã được phát trong phần thống kê ứng dụng. File này có các biến sau:

id	số thứ tự
age	tuổi
Gender	1: Nam, 2:nữ
Jobcat (vị trí công việc)	1: nhân viên, 2: trưởng nhóm/tổ trưởng/Quản đốc , 3:quản lý
educ	số năm đi học
prevexp	số tháng đã đi làm
salary	tiền lương hiện tại trong một năm (USD)
salbegin	tiền lương khởi điểm trong một năm (USD)

a. Bạn hãy tạo biến giả male: với quy ước 1: nam, 0 nữ; và biến female: 1 nữ, 0: nam ; từ biến jobcat, hãy tạo ra các biến giả phản ánh các vị trí công việc: clerical (1:nhân viên, 0 không phải nhân viên); teamleader (1: trưởng nhóm, 0: không phải là trưởng nhóm) ; manager (1:quản lý ; 0: không phải là quản lý). educmale là tích của 2 biến educ và male ; Từ đó, ước lượng mô hình sau:

$$\text{Salary} = \beta_1 + \beta_2 \text{educ} + \beta_3 \text{male} + \beta_4 \text{educmale} + \beta_5 \text{teamleader} + \beta_6 \text{manager} + u$$

b. Bạn hãy cho biết ý nghĩa kinh tế của hệ số hồi quy đứng trước biến manager, biến teamleader.

c. Theo bạn, hệ số hồi quy đứng trước biến educmale cho bạn biết điều gì?

d. Bạn hãy viết phương trình để ước tính tiền lương cho nam, phương trình để ước tính tiền lương cho nữ, và cho biết trong điều kiện cố định các yếu tố khác, trung bình mức chênh lệch tiền lương của nam so với nữ sẽ như thế nào?

f. Với mô hình ở câu a, bạn hãy dự đoán (dự báo điểm) mức lương trung bình của một người bằng bao nhiêu nếu biết người này có 16 năm đi học, ở chức vụ quản lý, là nam, có tuổi bằng 34, đã từng làm việc trong nghề là 5 năm, có một vợ hai con.

g. Bạn hãy ước lượng mô hình sau đây và cho biết những nhận xét của bạn từ kết quả tính toán:

$$\text{Salary} = \alpha_1 + \alpha_2 \text{educ} + \alpha_3 \text{female} + \alpha_4 \text{clerical} + \alpha_6 \text{manager} + v$$

Bài 2

Sử dụng dữ liệu baitap11c2.wfl để thực hiện các kết quả cần thiết và trả lời các câu hỏi.

Một ngân hàng ở khu vực Bình Phước đã thu thập dữ liệu của 60 khách hàng cá nhân đã từng vay vốn. Các khách hàng này chủ yếu là người trong một huyện, và họ mới di cư từ địa phương khác đến đây vài năm. Ngân hàng thu thập thông tin về tình hình trả được nợ của khách hàng khi hết hạn hợp đồng thông qua biến Y (1 – trả được nợ, 0 – không trả được nợ), học vấn (biến $hocvan$, đo lường bằng số năm đi học), thời gian nhập cư (biến $thoigian$, đo lường bằng tháng), tài sản thế chấp (biến $taisan$: 1 có tài sản thế chấp, 0 không có tài sản thế chấp). Người ta muốn xem xét khả năng trả được nợ của khách hàng được giải thích bởi những yếu tố nào và từ đó xây dựng mô hình xếp hạng tín dụng.

a. Bạn hãy ước lượng mô hình Logit cho biết khả năng trả được nợ của khách hàng phụ thuộc vào $hocvan$, $thoigian$, $taisan$; viết phương trình hồi quy mẫu.

b. Bạn hãy cho biết biến $hocvan$ có ý nghĩa thống kê ở độ tin cậy 95% hay không?

c. Bạn hãy cho biết biến $thoigian$ có ý nghĩa thống kê ở mức ý nghĩa 10% hay không?

d. Bạn hãy cho biết khả năng trả được nợ của khách hàng nam là cao hơn hay thấp hơn so với khách hàng nữ, hãy nhận xét về mối quan hệ giữa biến $hocvan$ và khả năng trả được nợ của khách hàng.

e. Tính tác động biên của $hocvan$ lên xác suất trả được nợ của khách hàng, giả sử bạn chọn xác suất ban đầu bằng tỷ lệ khách hàng trả được nợ

f. Bạn hãy dự đoán xác suất trả được nợ của một khách hàng có số năm đi học là 10, thời gian cư trú là 18 tháng, và có tài sản thế chấp. Nếu chỉ dựa vào mô hình mà bạn đã ước lượng ở câu a để xếp hạng tín dụng, bạn có nên cho khách hàng này vay không?

g. Theo bạn, độ chính xác của mô hình Logit trong dự đoán khả năng trả được nợ của khách hàng là bao nhiêu (Tính Count-R^2)?

h. Bạn hãy làm lại câu a, câu f nhưng với mô hình Probit

Bài 3.

Tác giả Nguyễn Thị Minh và các cộng sự (2011) sử dụng dữ liệu VHLSS 2008 và số liệu của VCCI đã nghiên cứu về một số yếu tố tác động đến nghèo đa chiều¹ (biến Y có giá trị bằng 1 khi một người là nghèo đa chiều, biến Y có giá trị bằng 0 khi một người không nghèo đa chiều) thông qua mô hình Logit. Các nhóm biến giải thích mà tác giả quan tâm:

Đặc trưng cá nhân, bao gồm: (1) các biến giả phản ánh 4 nhóm tuổi: 0-15, 15-55, 45-65, và >65; (2) Giới tính với 1 là nữ, 0 là nam.

¹ Nghèo đa chiều là một khái niệm đã được UNDP đề cập từ 1990, 1997 và vào năm 2010 UNDP đưa ra chỉ số nghèo đa chiều MPI (Multidimensional Poverty Index) trong báo cáo phát triển con người. Alkire & Foster (2007) đưa ra công thức tính chỉ số MPI thông qua 3 khía cạnh là giáo dục, y tế, và điều kiện sống. Ở Việt Nam, vào năm 2010, UNDP cũng đã tính toán MPI qua tám chỉ số con bao gồm: thu nhập, giáo dục, sức khỏe, an sinh xã hội, dịch vụ tiện ích, điều kiện sống, mức độ an toàn, và hòa nhập cộng đồng.

Đặc trưng của hộ gia đình, bao gồm: (1) biến giả Nông Thôn với 1 là nông thôn, 0 là thành thị; (2) Quy mô hộ; (3) biến giả trình độ học vấn của chủ hộ với 1 là đã có bằng tốt nghiệp THPT trở lên, 0 không phải như vậy; (4) Sức khỏe của chủ hộ, được đo lường thông qua số ngày chủ hộ bị ốm trong năm

Đặc trưng của địa phương, bao gồm: (1) GDP bình quân đầu người, (2) Cơ sở hạ tầng – đo lường thông qua phần trăm doanh nghiệp hài lòng với cơ sở hạ tầng của địa phương, (3) Tính minh bạch của chính quyền địa phương – đo lường thông qua phần trăm doanh nghiệp đánh giá tốt về tính minh bạch các tài liệu - kế hoạch của địa phương.

Kết quả (trích) dưới đây phân tích cho những người không phải chủ hộ

	Hệ số	Sai số chuẩn
Hạ tầng	-0.314	0.019
Minh bạch	-0.812	0.122
Nông thôn	1.173	0.055
GDP bình quân đầu người	-0.008	0.005
Trình độ học vấn của chủ hộ	-1.188	0.049
Sức khỏe của chủ hộ	0.003	0.001
Quy mô hộ	0.048	0.011
Giới tính (1: nữ, 0: nam)	0.337	0.044
Nhóm tuổi 0-15	-1.356	0.079
Nhóm tuổi 15-45	-0.296	0.075
Nhóm tuổi 45-65	-0.074	0.084
Hằng số	0.768	0.221

Ghi chú: Y (1: nghèo đa chiều, 0: không nghèo đa chiều)

- Theo bạn, Xác suất nghèo đa chiều có khác biệt giữa nông thôn và thành thị; giữa các mức học vấn (xét ở mức ý nghĩa 5%)?
- Bạn có nhận xét gì về mối liên hệ giữa tính minh bạch của chính quyền địa phương và xác suất nghèo đa chiều?
- Bạn có nhận xét gì về mối liên hệ giữa các nhóm tuổi và xác suất nghèo đa chiều?